OPEN ACCESS Freely available online

PLoS one

A Higher-Order Generalized Singular Value Decomposition for Comparison of Global mRNA Expression from Multiple Organisms

Sri Priya Ponnapalli[1], Michael A. Saunders[2], Charles F. Van Loan[3], Orly Alter[4]*

1 Department of Electrical and Computer Engineering, University of Texas at Austin, Texas, United States of America, **2** Department of Management Science and Engineering, Stanford University, Stanford, California, United States of America, **3** Department of Computer Science, Cornell University, Ithaca, New York, United States of America, **4** Scientific Computing and Imaging (SCI) Institute and Departments of Bioengineering and Human Genetics, University of Utah, Salt Lake City, Utah, United States of America

Abstract

The number of high-dimensional datasets recording multiple aspects of a single phenomenon is increasing in many areas of science, accompanied by a need for mathematical frameworks that can compare multiple large-scale matrices with different row dimensions. The only such framework to date, the generalized singular value decomposition (GSVD), is limited to two matrices. We mathematically define a higher-order GSVD (HO GSVD) for $N \geq 2$ matrices $D_i \in \mathbb{R}^{m_i \times n}$, each with full column rank. Each matrix is exactly factored as $D_i = U_i \Sigma_i V^T$, where V, identical in all factorizations, is obtained from the eigensystem $SV = V\Lambda$ of the arithmetic mean S of all pairwise quotients $A_i A_j^{-1}$ of the matrices $A_i = D_i^T D_i$, $i \neq j$. We prove that this decomposition extends to higher orders almost all of the mathematical properties of the GSVD. The matrix S is nondefective with V and Λ real. Its eigenvalues satisfy $\lambda_k \geq 1$. Equality holds if and only if the corresponding eigenvector v_k is a right basis vector of equal significance in all matrices D_i and D_j, that is $\sigma_{i,k}/\sigma_{j,k} = 1$ for all i and j, and the corresponding left basis vector $u_{i,k}$ is orthogonal to all other vectors in U_i for all i. The eigenvalues $\lambda_k = 1$, therefore, define the "common HO GSVD subspace." We illustrate the HO GSVD with a comparison of genome-scale cell-cycle mRNA expression from *S. pombe*, *S. cerevisiae* and human. Unlike existing algorithms, a mapping among the genes of these disparate organisms is not required. We find that the approximately common HO GSVD subspace represents the cell-cycle mRNA expression oscillations, which are similar among the datasets. Simultaneous reconstruction in the common subspace, therefore, removes the experimental artifacts, which are dissimilar, from the datasets. In the simultaneous sequence-independent classification of the genes of the three organisms in this common subspace, genes of highly conserved sequences but significantly different cell-cycle peak times are correctly classified.

Citation: Ponnapalli SP, Saunders MA, Van Loan CF, Alter O (2011) A Higher-Order Generalized Singular Value Decomposition for Comparison of Global mRNA Expression from Multiple Organisms. PLoS ONE 6(12): e28072. doi:10.1371/journal.pone.0028072

Editor: Dongxiao Zhu, Wayne State University, United States of America

Received September 22, 2011; **Accepted** October 31, 2011; **Published** December 22, 2011

Funding: This research was supported by Office of Naval Research Grant N00014-02-1-0076 (to MAS), National Science Foundation Grant DMS-1016284 (to CFVL), as well as the Utah Science Technology and Research (USTAR) Initiative, National Human Genome Research Institute R01 Grant HG-004302 and National Science Foundation CAREER Award DMS-0847173 (to OA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: orly@sci.utah.edu

Introduction

In many areas of science, especially in biotechnology, the number of high-dimensional datasets recording multiple aspects of a single phenomenon is increasing. This is accompanied by a fundamental need for mathematical frameworks that can compare multiple large-scale matrices with different row dimensions. For example, comparative analyses of global mRNA expression from multiple model organisms promise to enhance fundamental understanding of the universality and specialization of molecular biological mechanisms, and may prove useful in medical diagnosis, treatment and drug design [1]. Existing algorithms limit analyses to subsets of homologous genes among the different organisms, effectively introducing into the analysis the assumption that sequence and functional similarities are equivalent (e.g., [2]). However, it is well known that this assumption does not always hold, for example, in cases of nonorthologous gene displacement, when nonorthologous proteins in different organisms fulfill the same function [3]. For sequence-independent comparisons, mathematical frameworks are required that can distinguish and separate the similar from the dissimilar among multiple large-scale datasets tabulated as matrices with different row dimensions, corresponding to the different sets of genes of the different organisms. The only such framework to date, the generalized singular value decomposition (GSVD) [4–7], is limited to two matrices.

It was shown that the GSVD provides a mathematical framework for sequence-independent comparative modeling of DNA microarray data from two organisms, where the mathematical variables and operations represent biological reality [7,8]. The variables, significant subspaces that are common to both or exclusive to either one of the datasets, correlate with cellular programs that are conserved in both or unique to either one of the organisms, respectively. The operation of reconstruction in the

subspaces common to both datasets outlines the biological similarity in the regulation of the cellular programs that are conserved across the species. Reconstruction in the common and exclusive subspaces of either dataset outlines the differential regulation of the conserved relative to the unique programs in the corresponding organism. Recent experimental results [9] verify a computationally predicted genome-wide mode of regulation that correlates DNA replication origin activity with mRNA expression [10,11], demonstrating that GSVD modeling of DNA microarray data can be used to correctly predict previously unknown cellular mechanisms.

We now define a higher-order GSVD (HO GSVD) for the comparison of $N \geq 2$ datasets. The datasets are tabulated as N real matrices $D_i \in \mathbb{R}^{m_i \times n}$, each with full column rank, with different row dimensions and the same column dimension, where there exists a one-to-one mapping among the columns of the matrices. Like the GSVD, the HO GSVD is an exact decomposition, i.e., each matrix is exactly factored as $D_i = U_i \Sigma_i V^T$, where the columns of U_i and V have unit length and are the left and right basis vectors respectively, and each Σ_i is diagonal and positive definite. Like the GSVD, the matrix V is identical in all factorizations. In our HO GSVD, the matrix V is obtained from the eigensystem $SV = V\Lambda$ of the arithmetic mean S of all pairwise quotients $A_i A_j^{-1}$ of the matrices $A_i = D_i^T D_i$, or equivalently of all $S_{ij} = \frac{1}{2}(A_i A_j^{-1} + A_j A_i^{-1})$, $i \neq j$.

To clarify our choice of S, we note that in the GSVD, defined by Van Loan [5], the matrix V can be formed from the eigenvectors of the unbalanced quotient $A_1 A_2^{-1}$ (Section 1 in Appendix S1). We observe that this V can also be formed from the eigenvectors of the balanced arithmetic mean $S_{12} = \frac{1}{2}(A_1 A_2^{-1} + A_2 A_1^{-1})$. We prove that in the case of $N = 2$, our definition of V by using the eigensystem of $S \equiv S_{12} = \frac{1}{2}(A_1 A_2^{-1} + A_2 A_1^{-1})$ leads algebraically to the GSVD (Theorems S1–S5 in Appendix S1), and therefore, as Paige and Saunders showed [6], can be computed in a stable way. We also note that in the GSVD, the matrix V does not depend upon the ordering of the matrices D_1 and D_2. Therefore, we define our HO GSVD for $N \geq 2$ matrices by using the balanced arithmetic mean S of all pairwise arithmetic means S_{ij}, each of which defines the GSVD of the corresponding pair of matrices D_i and D_j, noting that S does not depend upon the ordering of the matrices D_i and D_j.

We prove that S is nondefective (it has n independent eigenvectors), and that its eigensystem is real (Theorem 1). We prove that the eigenvalues of S satisfy $\lambda_k \geq 1$ (Theorem 2). As in our GSVD comparison of two matrices [7], we interpret the kth diagonal of $\Sigma_i = \mathrm{diag}(\sigma_{i,k})$ in the factorization of the i th matrix D_i as indicating the significance of the kth right basis vector v_k in D_i in terms of the overall information that v_k captures in D_i. The ratio $\sigma_{i,k}/\sigma_{j,k}$ indicates the significance of v_k in D_i relative to its significance in D_j. We prove that an eigenvalue of S satisfies $\lambda_k = 1$ if and only if the corresponding eigenvector v_k is a right basis vector of equal significance in all D_i and D_j, that is, $\sigma_{i,k}/\sigma_{j,k} = 1$ for all i and j, and the corresponding left basis vector $u_{i,k}$ is orthonormal to all other vectors in U_i for all i. We therefore mathematically define, in analogy with the GSVD, the "common HO GSVD subspace" of the $N \geq 2$ matrices to be the subspace spanned by the right basis vectors v_k that correspond to the $\lambda_k = 1$ eigenvalues of S (Theorem 3). We also show that each of the right basis vectors $\{v_k\}$ that span the common HO GSVD subspace is a generalized singular vector of all pairwise GSVD factorizations of the matrices D_i and D_j with equal corresponding generalized singular values for all i and j (Corollary 1).

Recent research showed that several higher-order generalizations are possible for a given matrix decomposition, each preserving some but not all of the properties of the matrix decomposition [12–14] (see also Theorem S6 and Conjecture S1 in Appendix S1). Our new HO GSVD extends to higher orders all of the mathematical properties of the GSVD except for complete column-wise orthogonality of the left basis vectors that form the matrix U_i for all i, i.e., in each factorization.

We illustrate the HO GSVD with a comparison of cell-cycle mRNA expression from *S. pombe* [15,16], *S. cerevisiae* [17] and human [18]. Unlike existing algorithms, a mapping among the genes of these disparate organisms is not required (Section 2 in Appendix S1). We find that the common HO GSVD subspace represents the cell-cycle mRNA expression oscillations, which are similar among the datasets. Simultaneous reconstruction in this common subspace, therefore, removes the experimental artifacts, which are dissimilar, from the datasets. Simultaneous sequence-independent classification of the genes of the three organisms in the common subspace is in agreement with previous classifications into cell-cycle phases [19]. Notably, genes of highly conserved sequences across the three organisms [20,21] but significantly different cell-cycle peak times, such as genes from the ABC transporter superfamily [22–28], phospholipase B-encoding genes [29,30] and even the B cyclin-encoding genes [31,32], are correctly classified.

Methods

HO GSVD Construction

Suppose we have a set of N real matrices $D_i \in \mathbb{R}^{m_i \times n}$ each with full column rank. We define a HO GSVD of these N matrices as

$$\begin{aligned} D_1 &= U_1 \Sigma_1 V^T, \\ D_2 &= U_2 \Sigma_2 V^T, \\ &\vdots \\ D_N &= U_N \Sigma_N V^T, \end{aligned} \tag{1}$$

where each $U_i \in \mathbb{R}^{m_i \times n}$ is composed of normalized left basis vectors, each $\Sigma_i = \mathrm{diag}(\sigma_{i,k}) \in \mathbb{R}^{n \times n}$ is diagonal with $\sigma_{i,k} > 0$, and V, identical in all matrix factorizations, is composed of normalized right basis vectors. As in the GSVD comparison of global mRNA expression from two organisms [7], in the HO GSVD comparison of global mRNA expression from $N \geq 2$ organisms, the shared right basis vectors v_k of Equation (1) are the "genelets" and the N sets of left basis vectors $u_{i,k}$ are the N sets of "arraylets" (Figure 1 and Section 2 in Appendix S1). We obtain V from the eigensystem of S, the arithmetic mean of all pairwise quotients $A_i A_j^{-1}$ of the matrices $A_i = D_i^T D_i$, or equivalently of all $S_{ij} = \frac{1}{2}(A_i A_j^{-1} + A_j A_i^{-1})$, $i \neq j$:

$$\begin{aligned} S &\equiv \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j>i}^{N} \left(A_i A_j^{-1} + A_j A_i^{-1} \right) \\ &= \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j>i}^{N} S_{ij}, \end{aligned} \tag{2}$$

$$\begin{aligned} &SV = V\Lambda, \\ &V \equiv (v_1 \ldots v_n), \quad \Lambda = \mathrm{diag}(\lambda_k), \end{aligned}$$

with $\|v_k\| = 1$. We prove that S is nondefective, i.e., S has n independent eigenvectors, and that its eigenvectors V and eigenvalues Λ are real (Theorem 1). We prove that the eigenvalues of S satisfy $\lambda_k \geq 1$ (Theorem 2).

Given V, we compute matrices B_i by solving N linear systems:

$$\begin{aligned} V B_i^T &= D_i^T, \\ B_i &\equiv (b_{i,1} \ldots b_{i,n}), \quad i = 1, \ldots, N, \end{aligned} \tag{3}$$

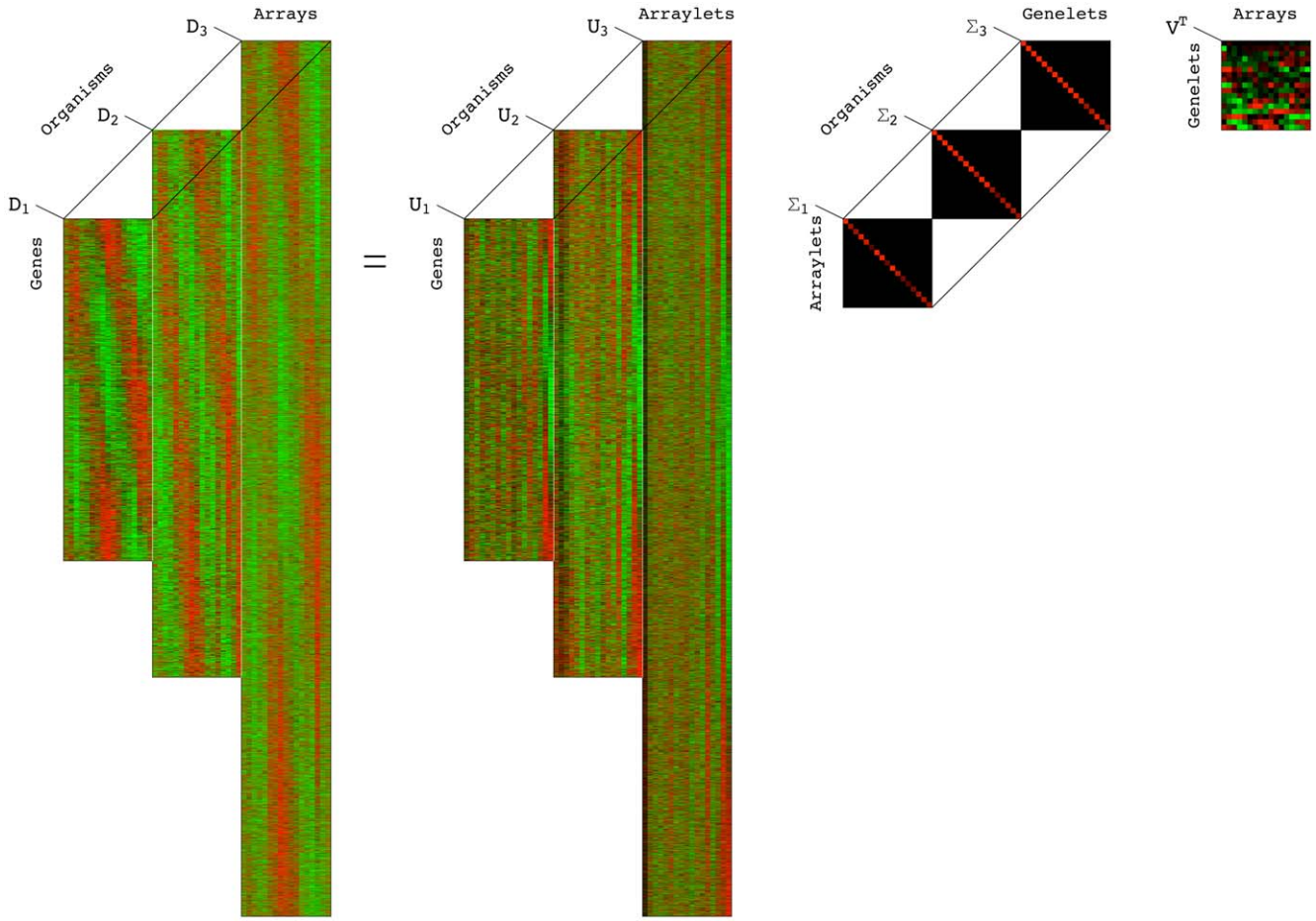

Figure 1. Higher-order generalized singular value decomposition (HO GSVD). In this raster display of Equation (1) with overexpression (red), no change in expression (black), and underexpression (green) centered at gene- and array-invariant expression, the *S. pombe*, S. cerevisiae and human global mRNA expression datasets are tabulated as organism-specific genes × 17-arrays matrices D_1, D_2 and D_3. The underlying assumption is that there exists a one-to-one mapping among the 17 columns of the three matrices but not necessarily among their rows. These matrices are transformed to the reduced diagonalized matrices Σ_1, Σ_2 and Σ_3, each of 17-"arraylets," i.e., left basis vectors × 17-"genelets," i.e., right basis vectors, by using the organism-specific genes × 17-arraylets transformation matrices U_1, U_2 and U_3 and the shared 17-genelets × 17-arrays transformation matrix V^T. We prove that with our particular V of Equations (2)–(4), this decomposition extends to higher orders all of the mathematical properties of the GSVD except for complete column-wise orthogonality of the arraylets, i.e., left basis vectors that form the matrices U_1, U_2 and U_3. We therefore mathematically define, in analogy with the GSVD, the "common HO GSVD subspace" of the $N=3$ matrices to be the subspace spanned by the genelets, i.e., right basis vectors v_k that correspond to higher-order generalized singular values that are equal, $\sigma_{1,k}=\sigma_{2,k}=\sigma_{3,k}$, where, as we prove, the corresponding arraylets, i.e., the left basis vectors $u_{1,k}$, $u_{2,k}$ and $u_{3,k}$, are orthonormal to all other arraylets in U_1, U_2 and U_3. We show that like the GSVD for two organisms [7], the HO GSVD provides a sequence-independent comparative mathematical framework for datasets from more than two organisms, where the mathematical variables and operations represent biological reality: Genelets of common significance in the multiple datasets, and the corresponding arraylets, represent cell-cycle checkpoints or transitions from one phase to the next, common to *S. pombe*, *S. cerevisiae* and human. Simultaneous reconstruction and classification of the three datasets in the common subspace that these patterns span outline the biological similarity in the regulation of their cell-cycle programs. Notably, genes of significantly different cell-cycle peak times [19] but highly conserved sequences [20,21] are correctly classified.
doi:10.1371/journal.pone.0028072.g001

and we construct Σ_i and $U_i=(u_{i,1} \quad \ldots \quad u_{i,n})$ by normalizing the columns of B_i:

$$\begin{aligned} \sigma_{i,k} &= \|b_{i,k}\|, \\ \Sigma_i &= \mathrm{diag}(\sigma_{i,k}), \\ B_i &= U_i\Sigma_i. \end{aligned} \tag{4}$$

HO GSVD Interpretation

In this construction, the rows of each of the N matrices D_i are superpositions of the same right basis vectors, the columns of V (Figures S1 and S2 and Section 1 in Appendix S1). As in our GSVD comparison of two matrices, we interpret the kth diagonals of Σ_i, the "higher-order generalized singular value set" $\{\sigma_{i,k}\}$, as indicating the significance of the kth right basis vector v_k in the matrices D_i, and reflecting the overall information that v_k captures in each D_i respectively. The ratio $\sigma_{i,k}/\sigma_{j,k}$ indicates the significance of v_k in D_i relative to its significance in D_j. A ratio of $\sigma_{i,k}/\sigma_{j,k}=1$ for all i and j corresponds to a right basis vector v_k of equal significance in all N matrices D_i. GSVD comparisons of two matrices showed that right basis vectors of approximately equal significance in the two matrices reflect themes that are common to both matrices under comparison [7]. A ratio of $\sigma_{i,k}/\sigma_{j,k} \ll 1$ indicates a basis vector v_k of almost negligible

significance in D_i relative to its significance in D_j. GSVD comparisons of two matrices showed that right basis vectors of negligible significance in one matrix reflect themes that are exclusive to the other matrix.

We prove that an eigenvalue of S satisfies $\lambda_k = 1$ if and only if the corresponding eigenvector v_k is a right basis vector of equal significance in all D_i and D_j, that is, $\sigma_{i,k}/\sigma_{j,k} = 1$ for all i and j, and the corresponding left basis vector $u_{i,k}$ is orthonormal to all other vectors in U_i for all i. We therefore mathematically define, in analogy with the GSVD, the "common HO GSVD subspace" of the $N \geq 2$ matrices to be the subspace spanned by the right basis vectors $\{v_k\}$ corresponding to the eigenvalues of S that satisfy $\lambda_k = 1$ (Theorem 3).

It follows that each of the right basis vectors $\{v_k\}$ that span the common HO GSVD subspace is a generalized singular vector of all pairwise GSVD factorizations of the matrices D_i and D_j with equal corresponding generalized singular values for all i and j (Corollary 1). Since the GSVD can be computed in a stable way [6], we note that the common HO GSVD subspace can also be computed in a stable way by computing all pairwise GSVD factorizations of the matrices D_i and D_j. This also suggests that it may be possible to formulate the HO GSVD as a solution to an optimization problem, in analogy with existing variational formulations of the GSVD [33]. Such a formulation may lead to a stable numerical algorithm for computing the HO GSVD, and possibly also to a higher-order general Gauss-Markov linear statistical model [34–36].

We show, in a comparison of $N = 3$ matrices, that the approximately common HO GSVD subspace of these three matrices reflects a theme that is common to the three matrices under comparison (Section 2).

HO GSVD Mathematical Properties

Theorem 1. *S is nondefective (it has n independent eigenvectors) and its eigensystem is real.*

Proof. From Equation (2) it follows that

$$\begin{aligned} S &= \frac{1}{N(N-1)}(H - NI), \\ H &= \left(\sum_{i=1}^{N} A_i\right)\left(\sum_{j=1}^{N} A_j^{-1}\right), \end{aligned} \tag{5}$$

and the eigenvectors of S equal the eigenvectors of H.

Let the SVD of the matrices D_i appended along the n-columns axis be

$$\begin{aligned} \begin{bmatrix} D_1 \\ \vdots \\ D_N \end{bmatrix} &= \begin{bmatrix} \hat{U}_1 \\ \vdots \\ \hat{U}_N \end{bmatrix} \hat{\Sigma} \hat{V}^T, \\ \sum_{i=1}^{N} \hat{U}_i^T \hat{U}_i &= I. \end{aligned} \tag{6}$$

Since the matrices D_i are real and with full column rank, it follows from the SVD of $\hat{U}_i$ that the symmetric matrices $\hat{U}_i^T \hat{U}_i$ are real and positive definite, and their inverses exist. It then follows from Equations (5) and (6) that H is similar to $\hat{H}$,

$$\begin{aligned} H &= \hat{V}\hat{\Sigma}\hat{H}\hat{\Sigma}^{-1}\hat{V}^T, \\ \hat{H} &= \sum_{j=1}^{N} (\hat{U}_j^T \hat{U}_j)^{-1}, \end{aligned} \tag{7}$$

and the eigenvalues of H equal the eigenvalues of $\hat{H}$.

A sum of real, symmetric and positive definite matrices, $\hat{H}$ is also real, symmetric and positive definite; therefore, its eigensystem

$$\hat{Y}^T \hat{H} \hat{Y} = \mathrm{diag}(\mu_k) \tag{8}$$

is real with $\hat{Y}$ orthogonal and $\mu_k > 0$. Without loss of generality let $\hat{Y}$ be orthonormal, such that $\|\hat{y}_k\| = 1$. It follows from the similarity of H with $\hat{H}$ that the eigensystem of H can be written as $V^{-1} H V = \mathrm{diag}(\mu_k)$, with the real and nonsingular $V = (\hat{V}\hat{\Sigma}\hat{Y})\hat{W}^{-1}$, where $\hat{W} = \mathrm{diag}(\hat{w}_k)$ and $\hat{w}_k = \|\hat{V}\hat{\Sigma}\hat{y}_k\|$ such that $\|v_k\| = 1$ for all k.

Thus, from Equation (5), S is nondefective with real eigenvectors V. Also, the eigenvalues of S satisfy

$$\lambda_k = \frac{1}{N(N-1)}(\mu_k - N), \tag{9}$$

where $\mu_k > 0$ are the eigenvalues of H and $\hat{H}$. Thus, the eigenvalues of S are real. □

Theorem 2. *The eigenvalues of S satisfy $\lambda_k \geq 1$.*

Proof. Following Equation (9), asserting that the eigenvalues of S satisfy $\lambda_k \geq 1$ is equivalent to asserting that the eigenvalues of $\hat{H}$ satisfy $\mu_k \geq N^2$.

From Equations (6) and (7), the eigenvalues of $\hat{H}$ satisfy

$$\mu_k \geq \min_x \sum_{j=1}^{N} [x^T (\hat{U}_j^T \hat{U}_j) x]^{-1}, \tag{10}$$

under the constraint that

$$\sum_{j=1}^{N} x^T (\hat{U}_j^T \hat{U}_j) x = 1, \tag{11}$$

where x is a real unit vector, and where it follows from the Cauchy-Schwarz inequality [37] (see also [4,34,38]) for the real nonzero vectors $(\hat{U}_j^T \hat{U}_j)x$ and $(\hat{U}_j^T \hat{U}_j)^{-1}x$ that for all j

$$x^T (\hat{U}_j^T \hat{U}_j)^{-1} x \geq [x^T (\hat{U}_j^T \hat{U}_j) x]^{-1}. \tag{12}$$

With the constraint of Equation (11), which requires the sum of the N positive numbers $x^T(\hat{U}_j^T \hat{U}_j)^{-1}x$ to equal one, the lower bound on the eigenvalues of $\hat{H}$ in Equation (10) is at its minimum when the sum of the inverses of these numbers is at its minimum, that is, when the numbers equal

$$x^T (\hat{U}_i^T \hat{U}_i) x = x^T (\hat{U}_j^T \hat{U}_j) x = N^{-1} \tag{13}$$

for all i and j. Thus, the eigenvalues of $\hat{H}$ satisfy $\mu_k \geq N^2$. □

Theorem 3. ***The common HO GSVD subspace.*** *An eigenvalue of S satisfies $\lambda_k = 1$ if and only if the corresponding eigenvector v_k is a right basis vector of equal significance in all D_i and D_j, that is, $\sigma_{i,k}/\sigma_{j,k} = 1$ for all i and j, and the corresponding left basis vector $u_{i,k}$ is orthonormal to all other vectors in U_i for all i. The "common HO GSVD subspace" of the $N \geq 2$ matrices is, therefore, the subspace spanned by the right basis vectors $\{v_k\}$ corresponding to the eigenvalues of S that satisfy $\lambda_k = 1$.*

Proof. Without loss of generality, let $k = n$. From Equation (12) and the Cauchy-Schwarz inequality, an eigenvalue of $\hat{H}$ equals its minimum lower bound $\mu_n = N^2$ if and only if the corresponding

eigenvector $\hat{y}_n$ is also an eigenvector of $\hat{U}_i^T \hat{U}_i$ for all i [37], where, from Equation (13), the corresponding eigenvalue equals N^{-1},

$$(\hat{U}_i^T \hat{U}_i)\hat{Y} = [(\hat{U}_i^T \hat{U}_i)\hat{y}_1 \quad (\hat{U}_i^T \hat{U}_i)\hat{y}_2 \quad \dots \quad N^{-1}\hat{y}_n]. \qquad (14)$$

Given the eigenvectors $V = (\hat{V}\hat{\Sigma}\hat{Y})\hat{W}^{-1}$ of S, we solve Equation (3) for each $D_i = \hat{U}_i\hat{\Sigma}\hat{V}^T$ of Equation (6), and obtain

$$\begin{aligned} U_i\Sigma_i &= B_i = D_i V^{-T} \\ &= \hat{U}_i\hat{Y}\hat{W}. \end{aligned} \qquad (15)$$

Following Equations (14) and (15), where $v_n = \hat{w}_n^{-1}\hat{V}\hat{\Sigma}\hat{y}_n$ corresponds to a minimum eigenvalue $\lambda_n = 1$, and since $\hat{Y}$ is orthonormal, we obtain

$$\begin{aligned} &\hat{W}^{-1}\Sigma_i(U_i^T U_i)\Sigma_i\hat{W}^{-1} = \hat{Y}^T(\hat{U}_i^T\hat{U}_i)\hat{Y} \\ &= \begin{bmatrix} \hat{y}_1^T(\hat{U}_i^T\hat{U}_i)\hat{y}_1 & \hat{y}_2^T(\hat{U}_i^T\hat{U}_i)\hat{y}_1 & \dots & 0 \\ \hat{y}_1^T(\hat{U}_i^T\hat{U}_i)\hat{y}_2 & \hat{y}_2^T(\hat{U}_i^T\hat{U}_i)\hat{y}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & N^{-1} \end{bmatrix}, \end{aligned} \qquad (16)$$

with zeroes in the nth row and the nth column of the matrix above everywhere except for the diagonal element. Thus, an eigenvalue of S satisfies $\lambda_n = 1$ if and only if the corresponding left basis vectors $u_{i,n}$ are orthonormal to all other vectors in U_i.

The corresponding higher-order generalized singular values are $\sigma_{i,n} = N^{-1/2}\hat{w}_n > 0$. Thus $\sigma_{i,n}/\sigma_{j,n} = 1$ for all i and j, and the corresponding right basis vector v_n is of equal significance in all matrices D_i and D_j. □

Corollary 1. *An eigenvalue of S satisfies $\lambda_k = 1$ if and only if the corresponding right basis vector v_k is a generalized singular vector of all pairwise GSVD factorizations of the matrices D_i and D_j with equal corresponding generalized singular values for all i and j.*

Proof. From Equations (12) and (13), and since the pairwise quotients $A_iA_j^{-1}$ are similar to $(U_i^T U_i)(U_j^T U_j)^{-1}$ with the similarity transformation of $\hat{V}\hat{\Sigma}$ for all i and j, it follows that an eigenvalue of S satisfies $\lambda_k = 1$ if and only if the corresponding right basis vector $v_k = \hat{w}_k^{-1}\hat{V}\hat{\Sigma}\hat{y}_k$ is also an eigenvector of each of the pairwise quotients $A_iA_j^{-1}$ of the matrices $A_i = D_i^T D_i$ with equal corresponding eigenvalues, or equivalently of all S_{ij} with all eigenvalues at their minimum of one,

$$S_{ij}v_k = \frac{1}{2}(A_iA_j^{-1} + A_iA_j^{-1})v_k = v_k. \qquad (17)$$

We prove (Theorems S1–S5 in Appendix S1) that in the case of $N = 2$ matrices our definition of V by using the eigensystem of S_{ij} leads algebraically to the GSVD, where an eigenvalue of S_{ij} equals its minimum of one if and only if the two corresponding generalized singular values are equal, such that the corresponding generalized singular vector v_k is of equal significance in both matrices D_i and D_j. Thus, it follows that each of the right basis vectors $\{v_k\}$ that span the common HO GSVD subspace is a generalized singular vector of all pairwise GSVD factorizations of the matrices D_i and D_j with equal corresponding generalized singular values for all i and j. □

Note that since the GSVD can be computed in a stable way [6], the common HO GSVD subspace we define (Theorem 3) can also be computed in a stable way by computing all pairwise GSVD factorizations of the matrices D_i and D_j (Corollary 1). It may also be possible to formulate the HO GSVD as a solution to an optimization problem, in analogy with existing variational formulations of the GSVD [33]. Such a formulation may lead to a stable numerical algorithm for computing the HO GSVD, and possibly also to a higher-order general Gauss-Markov linear statistical model [34–36].

Results

HO GSVD Comparison of Global mRNA Expression from Three Organisms

Consider now the HO GSVD comparative analysis of global mRNA expression datasets from the $N = 3$ organisms *S. pombe*, *S. cerevisiae* and human (Section 2.1 in Appendix S1, Mathematica Notebooks S1 and S2, and Datasets S1, S2 and S3). The datasets are tabulated as matrices of $n = 17$ columns each, corresponding to DNA microarray-measured mRNA expression from each organism at 17 time points equally spaced during approximately two cell-cycle periods. The underlying assumption is that there exists a one-to-one mapping among the 17 columns of the three matrices but not necessarily among their rows, which correspond to either $m_1 = 3167$-*S. pombe* genes, $m_2 = 4772$-*S. cerevisiae* genes or $m_3 = 13{,}068$-human genes. The HO GSVD of Equation (1) transforms the datasets from the organism-specific genes × 17-arrays spaces to the reduced spaces of the 17-"arraylets," i.e., left basis vactors × 17-"genelets," i.e., right basis vectors, where the datasets D_i are represented by the diagonal nonnegative matrices Σ_i, by using the organism-specific genes × 17-arraylets transformation matrices U_i and the one shared 17-genelets × 17-arrays transformation matrix V^T (Figure 1).

Following Theorem 3, the approximately common HO GSVD subspace of the three datasets is spanned by the five genelets $k = 13, \dots, 17$ that correspond to $1 \lesssim \lambda_k \lesssim 2$. We find that these five genelets are approximately equally significant with $\sigma_{1,k} : \sigma_{2,k} : \sigma_{3,k} \sim 1 : 1 : 1$ in the *S. pombe*, *S. cerevisiae* and human datasets, respectively (Figure 2 *a* and *b*). The five corresponding arraylets in each dataset are $\varepsilon = 0.33$-orthonormal to all other arraylets (Figure S3 in Appendix S1).

Common HO GSVD Subspace Represents Similar Cell-Cycle Oscillations

The expression variations across time of the five genelets that span the approximately common HO GSVD subspace fit normalized cosine functions of two periods, superimposed on time-invariant expression (Figure 2 *c* and *d*). Consistently, the corresponding organism-specific arraylets are enriched [39] in overexpressed or underexpressed organism-specific cell cycle-regulated genes, with 24 of the 30 *P*-values $< 10^{-8}$ (Table 1 and Section 2.2 in Appendix S1). For example, the three 17th arraylets, which correspond to the 0-phase 17th genelet, are enriched in overexpressed G2 *S. pombe* genes, G2/M and M/G1 *S. cerevisiae* genes and S and G2 human genes, respectively, representing the cell-cycle checkpoints in which the three cultures are initially synchronized.

Simultaneous sequence-independent reconstruction and classification of the three datasets in the common subspace outline cell-cycle progression in time and across the genes in the three organisms (Sections 2.3 and 2.4 in Appendix S1). Projecting the expression of the 17 arrays of either organism from the corresponding five-dimensional arraylets subspace onto the two-dimensional subspace that approximates it (Figure S4 in Appendix S1), ≥50% of the contributions of the arraylets add

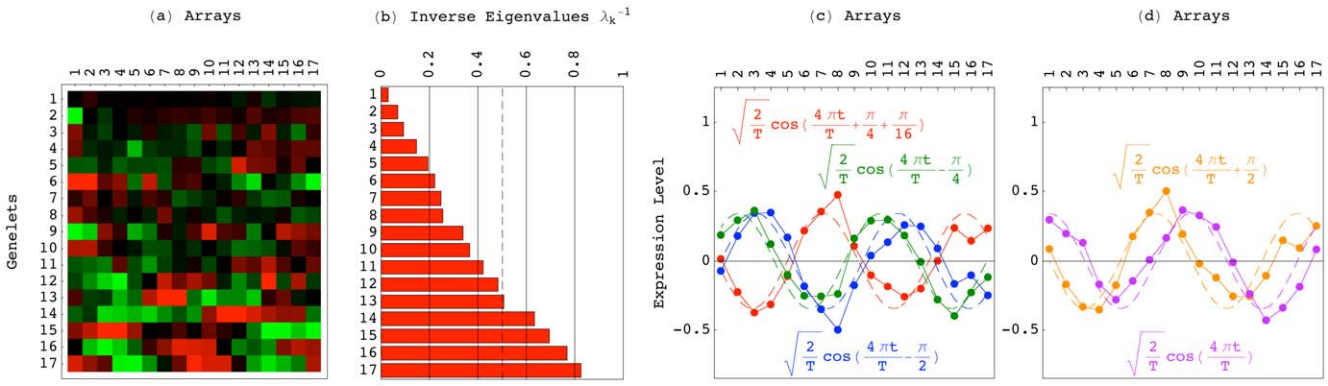

Figure 2. Genelets or right basis vectors. (*a*) Raster display of the expression of the 17 genelets, i.e., HO GSVD patterns of expression variation across time, with overexpression (red), no change in expression (black) and underexpression (green) around the array-, i.e., time-invariant expression. (*b*) Bar chart of the corresponding inverse eigenvalues λ_k^{-1}, showing that the 13th through the 17th genelets correspond to $1 \lesssim \lambda_k \lesssim 2$. (*c*) Line-joined graphs of the 13th (red), 14th (blue) and 15th (green) genelets in the two-dimensional subspace that approximates the five-dimensional HO GSVD subspace (Figure S4 and Section 2.4), normalized to zero average and unit variance. (*d*) Line-joined graphs of the projected 16th (orange) and 17th (violet) genelets in the two-dimensional subspace. The five genelets describe expression oscillations of two periods in the three time courses.
doi:10.1371/journal.pone.0028072.g002

up, rather than cancel out (Figure 3 *a–c*). In these two-dimensional subspaces, the angular order of the arrays of either organism describes cell-cycle progression in time through approximately two cell-cycle periods, from the initial cell-cycle phase and back to that initial phase twice. Projecting the expression of the genes, $\geq 50\%$ of the contributions of the five genelets add up in the overall expression of 343 of the 380 *S. pombe* genes classified as cell cycle-regulated, 554 of the 641 *S. cerevisiae* cell-cycle genes, and 632 of the 787 human cell-cycle genes (Figure 3 *d–f*). Simultaneous classification of the genes of either organism into cell-cycle phases according to their angular order in these two-dimensional subspaces is consistent with the classification of the arrays, and is in good agreement with the previous classifications of the genes (Figure 3 *g–i*). With all 3167 *S. pombe*, 4772 *S. cerevisiae* and 13,068 human genes sorted, the expression variations of the five arraylets from each organism approximately fit one-period cosines, with the initial phase of each arraylet (Figures S5, S6, S7 in Appendix S1) similar to that of its corresponding genelet (Figure 2). The global mRNA expression of each organism, reconstructed in the common HO GSVD subspace, approximately fits a traveling wave, oscillating across time and across the genes.

Table 1. Arraylets or left basis vectors.

		Overexpression		Underexpression	
Dataset	**Arraylet**	**Annotation**	***P*-value**	**Annotation**	***P*-value**
S. pombe	13	G2	2.4×10^{-10}	G1	1.0×10^{-15}
	14	M	2.2×10^{-21}	G2	1.3×10^{-9}
	15	M	4.1×10^{-13}	S	1.6×10^{-17}
	16	G2	5.2×10^{-18}	G1	1.2×10^{-26}
	17	G2	2.4×10^{-10}	S	5.3×10^{-35}
S. cerevisiae	13	S/G2	4.3×10^{-15}	M/G1	1.4×10^{-9}
	14	M/G1	4.9×10^{-26}	G2/M	2.2×10^{-12}
	15	G1	7.7×10^{-17}	S	1.3×10^{-8}
	16	G2/M	2.3×10^{-38}	G1	2.0×10^{-32}
	17	G2/M	2.3×10^{-41}	G1	2.6×10^{-40}
Human	13	G1/S	1.1×10^{-33}	G2	2.4×10^{-44}
	14	M/G1	5.7×10^{-3}	G2	4.7×10^{-2}
	15	G2	9.8×10^{-24}	None	1.4×10^{-1}
	16	G1/S	9.8×10^{-13}	G2	4.1×10^{-4}
	17	G2	9.3×10^{-33}	M/G1	2.7×10^{-2}

Probabilistic significance of the enrichment of the arraylets, i.e., HO GSVD patterns of expression variation across the *S. pombe*, *S. cerevisiae* and human genes, that span the common HO GSVD subspace in each dataset, in over- or underexpressed cell cycle-regulated genes. The *P*-value of each enrichment is calculated as described [39] (Section 2.2 in Appendix S1) assuming hypergeometric distribution of the annotations (Datasets S1, S2, S3) among the genes, including the $m = 100$ genes most over- or underexpressed in each arraylet.
doi:10.1371/journal.pone.0028072.t001

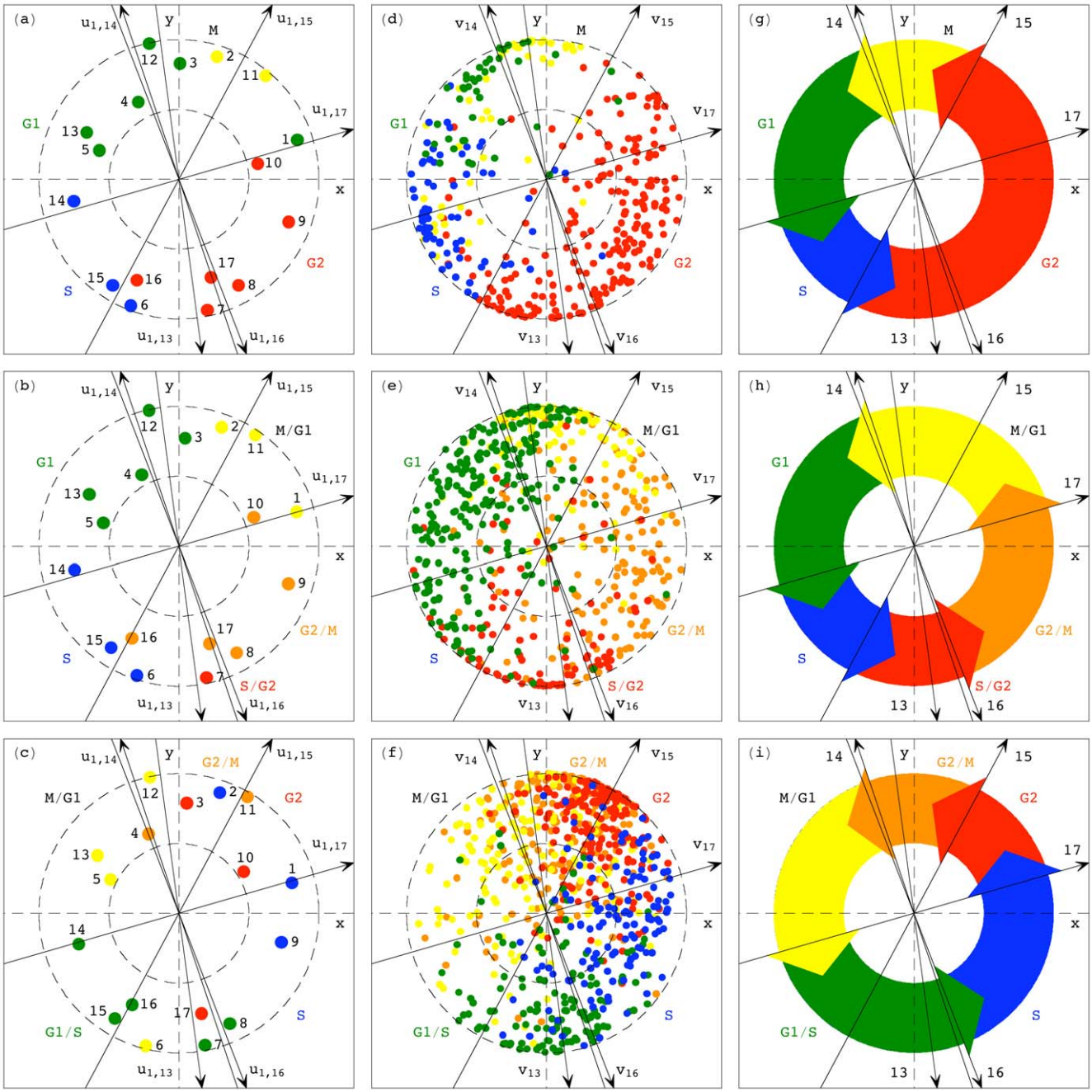

Figure 3. Common HO GSVD subspace represents similar cell-cycle oscillations. (*a–c*) *S. pombe*, *S. cerevisiae* and human array expression, projected from the five-dimensional common HO GSVD subspace onto the two-dimensional subspace that approximates it (Sections 2.3 and 2.4 in Appendix S1). The arrays are color-coded according to their previous cell-cycle classification [15–18]. The arrows describe the projections of the $k=13,\ldots,17$ arraylets of each dataset. The dashed unit and half-unit circles outline 100% and 50% of added-up (rather than canceled-out) contributions of these five arraylets to the overall projected expression. (*d–f*) Expression of 380, 641 and 787 cell cycle-regulated genes of *S. pombe*, *S. cerevisiae* and human, respectively, color-coded according to previous classifications. (*g–i*) The HO GSVD pictures of the *S. pombe*, *S. cerevisiae* and human cell-cycle programs. The arrows describe the projections of the $k=13,\ldots,17$ shared genelets and organism-specific arraylets that span the common HO GSVD subspace and represent cell-cycle checkpoints or transitions from one phase to the next.
doi:10.1371/journal.pone.0028072.g003

Note also that simultaneous reconstruction in the common HO GSVD subspace removes the experimental artifacts and batch effects, which are dissimilar, from the three datasets. Consider, for example, the second genelet. With $\sigma_{1,2} : \sigma_{2,2} : \sigma_{3,2} \sim 1 : 8 : 3$ in the *S. pombe*, *S. cerevisiae* and human datasets, respectively, this genelet is almost exclusive to the *S. cerevisiae* dataset. This genelet is anticorrelated with a time decaying pattern of expression (Figure 2*a*). Consistently, the corresponding *S. cerevisiae*-specific arraylet is enriched in underexpressed *S. cerevisiae* genes that were classified as up-regulated by the *S. cerevisiae* synchronizing agent, the α-factor pheromone, with the *P*-value $<10^{-46}$. Reconstruction in the common subspace effectively removes this *S. cerevisiae*-

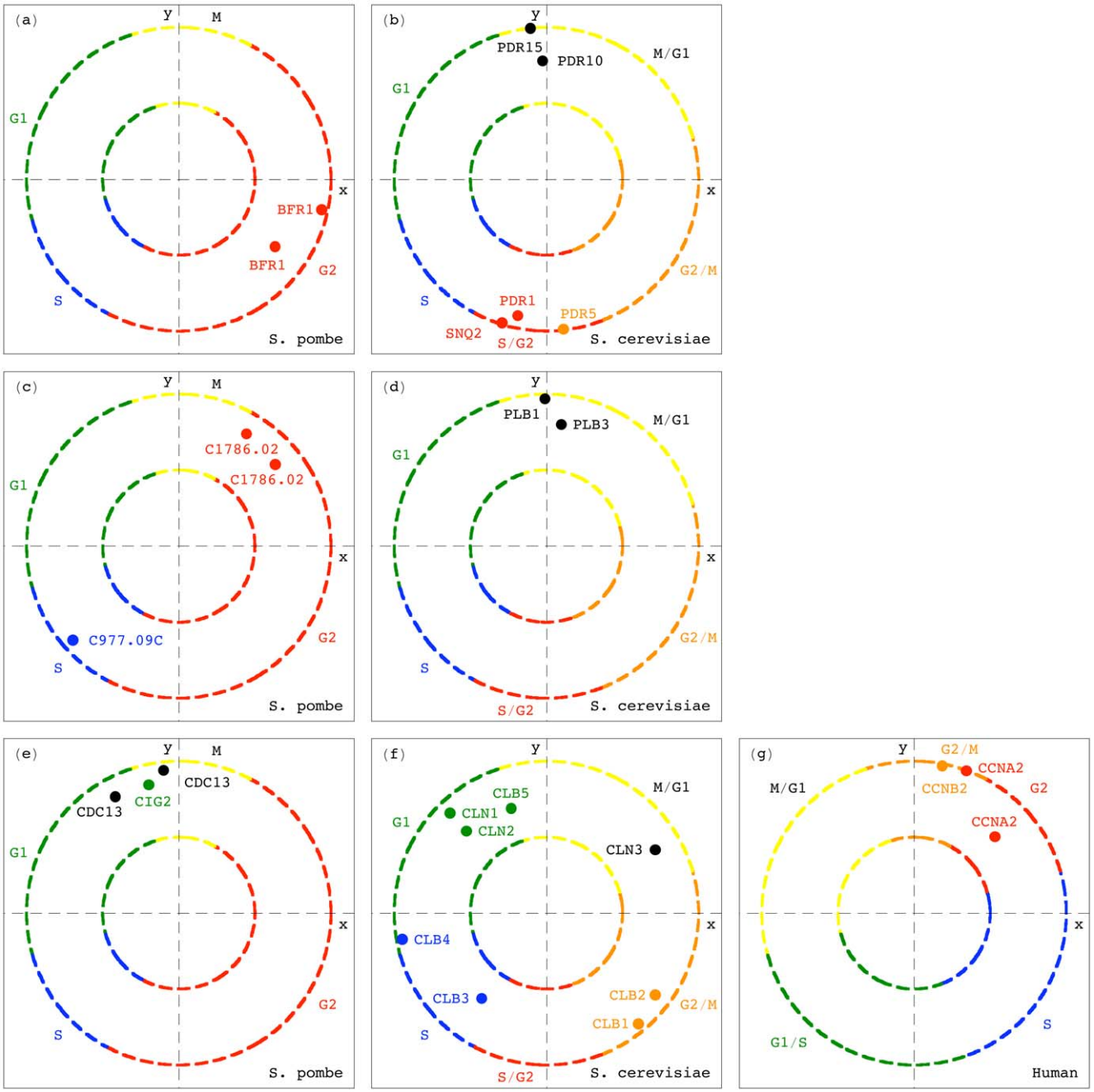

Figure 4. Simultaneous HO GSVD classification of homologous genes of different cell-cycle peak times. (*a*) The *S. pombe* gene *BFR1*, and (*b*) its closest *S. cerevisiae* homologs. (*c*) The *S. pombe* and (*d*) *S. cerevisiae* closest homologs of the *S. cerevisiae* gene *PLB1*. (*e*) The *S. pombe* cyclin-encoding gene *CIG2* and its closest *S. pombe*, (*f*) *S. cerevisiae* and (*g*) human homologs.
doi:10.1371/journal.pone.0028072.g004

approximately exclusive pattern of expression variation from the three datasets.

Simultaneous HO GSVD Classification of Homologous Genes of Different Cell-Cycle Peak Times

Notably, in the simultaneous sequence-independent classification of the genes of the three organisms in the common subspace, genes of significantly different cell-cycle peak times [19] but highly conserved sequences [20,21] are correctly classified (Section 2.5 in Appendix S1).

For example, consider the G2 *S. pombe* gene *BFR1* (Figure 4*a*), which belongs to the evolutionarily highly conserved ATP-binding cassette (ABC) transporter superfamily [22]. The closest homologs of *BFR1* in our *S. pombe*, *S. cerevisiae* and human datasets are the *S. cerevisiae* genes *SNQ2*, *PDR5*, *PDR15* and *PDR10* (Table S1*a* in Appendix S1). The expression of *SNQ2* and *PDR5* is known to peak at the S/G2 and G2/M cell-cycle phases, respectively [17]. However, sequence similarity does not imply similar cell-cycle peak times, and *PDR15* and *PDR10*, the closest homologs of *PDR5*, are induced during stationary phase [23], which has been

hypothesized to occur in G1, before the Cdc28-defined cell-cycle arrest [24]. Consistently, we find *PDR15* and *PDR10* at the M/G1 to G1 transition, antipodal to (i.e., half a cell-cycle period apart from) *SNQ2* and *PDR5*, which are projected onto S/G2 and G2/M, respectively (Figure 4*b*). We also find the transcription factor *PDR1* at S/G2, its known cell-cycle peak time, adjacent to *SNQ2* and *PDR5*, which it positively regulates and might be regulated by, and antipodal to *PDR15*, which it negatively regulates [25–28].

Another example is the *S. cerevisiae* phospholipase B-encoding gene *PLB1* [29], which peaks at the cell-cycle phase M/G1 [30]. Its closest homolog in our *S. cerevisiae* dataset, *PLB3*, also peaks at M/G1 [17] (Figure 4*d*). However, among the closest *S. pombe* and human homologs of *PLB1* (Table S1*b* in Appendix S1), we find the *S. pombe* genes *SPAC977.09c* and *SPAC1786.02*, which expressions peak at the almost antipodal *S. pombe* cell-cycle phases S and G2, respectively [19] (Figure 4*c*).

As a third example, consider the *S. pombe* G1 B-type cyclin-encoding gene *CIG2* [31,32] (Table S1*c* in Appendix S1). Its closest *S. pombe* homolog, *CDC13*, peaks at M [19] (Figure 4*e*). The closest human homologs of *CIG2*, the cyclins *CCNA2* and *CCNB2*, peak at G2 and G2/M, respectively (Figure 4*g*). However, while periodicity in mRNA abundance levels through the cell cycle is highly conserved among members of the cyclin family, the cell-cycle peak times are not necessarily conserved [1]: The closest homologs of *CIG2* in our *S. cerevisiae* dataset, are the G2/M promoter-encoding genes *CLB1,2* and *CLB3,4*, which expressions peak at G2/M and S respectively, and *CLB5*, which encodes a DNA synthesis promoter, and peaks at G1 (Figure 4*f*).

Discussion

We mathematically defined a higher-order GSVD (HO GSVD) for two or more large-scale matrices with different row dimensions and the same column dimension. We proved that our new HO GSVD extends to higher orders almost all of the mathematical properties of the GSVD: The eigenvalues of S are always greater than or equal to one, and an eigenvalue of one corresponds to a right basis vector of equal significance in all matrices, and to a left basis vector in each matrix factorization that is orthogonal to all other left basis vectors in that factorization. We therefore mathematically defined, in analogy with the GSVD, the common HO GSVD subspace of the $N \geq 2$ matrices to be the subspace spanned by the right basis vectors that correspond to the eigenvalues of S that equal one.

The only property that does not extend to higher orders in general is the complete column-wise orthogonality of the normalized left basis vectors in each factorization. Recent research showed that several higher-order generalizations are possible for a given matrix decomposition, each preserving some but not all of the properties of the matrix decomposition [12–14]. The HO GSVD has the interesting property of preserving the exactness and diagonality of the matrix GSVD and, in special cases, also partial or even complete column-wise orthogonality. That is, all N matrix factorizations in Equation (1) are exact, all N matrices Σ_i are diagonal, and when one or more of the eigenvalues of S equal one, the corresponding left basis vectors in each factorization are orthogonal to all other left basis vectors in that factorization.

The complete column-wise orthogonality of the matrix GSVD [5] enables its stable computation [6]. We showed that each of the right basis vectors that span the common HO GSVD subspace is a generalized singular vector of all pairwise GSVD factorizations of the matrices D_i and D_j with equal corresponding generalized singular values for all i and j. Since the GSVD can be computed in a stable way, the common HO GSVD subspace can also be computed in a stable way by computing all pairwise GSVD factorizations of the matrices D_i and D_j. That is, the common HO GSVD subspace exists also for N matrices D_i that are not all of full column rank. This also means that the common HO GSVD subspace can be formulated as a solution to an optimization problem, in analogy with existing variational formulations of the GSVD [33].

It would be ideal if our procedure reduced to the stable computation of the matrix GSVD when $N=2$. To achieve this ideal, we would need to find a procedure that allows a computation of the HO GSVD, not just the common HO GSVD subspace, for N matrices D_i that are not all of full column rank. A formulation of the HO GSVD, not just the common HO GSVD subspace, as a solution to an optimization problem may lead to a stable numerical algorithm for computing the HO GSVD. Such a formulation may also lead to a higher-order general Gauss-Markov linear statistical model [34–36].

It was shown that the GSVD provides a mathematical framework for sequence-independent comparative modeling of DNA microarray data from two organisms, where the mathematical variables and operations represent experimental or biological reality [7,8]. The variables, subspaces of significant patterns that are common to both or exclusive to either one of the datasets, correlate with cellular programs that are conserved in both or unique to either one of the organisms, respectively. The operation of reconstruction in the subspaces common to both datasets outlines the biological similarity in the regulation of the cellular programs that are conserved across the species. Reconstruction in the common and exclusive subspaces of either dataset outlines the differential regulation of the conserved relative to the unique programs in the corresponding organism. Recent experimental results [9] verify a computationally predicted genome-wide mode of regulation [10,11], and demonstrate that GSVD modeling of DNA microarray data can be used to correctly predict previously unknown cellular mechanisms.

Here we showed, comparing global cell-cycle mRNA expression from the three disparate organisms *S. pombe*, *S. cerevisiae* and human, that the HO GSVD provides a sequence-independent comparative framework for two or more genomic datasets, where the variables and operations represent biological reality. The approximately common HO GSVD subspace represents the cell-cycle mRNA expression oscillations, which are similar among the datasets. Simultaneous reconstruction in the common subspace removes the experimental artifacts, which are dissimilar, from the datasets. In the simultaneous sequence-independent classification of the genes of the three organisms in this common subspace, genes of highly conserved sequences but significantly different cell-cycle peak times are correctly classified.

Additional possible applications of our HO GSVD in biotechnology include comparison of multiple genomic datasets, each corresponding to (*i*) the same experiment repeated multiple times using different experimental protocols, to separate the biological signal that is similar in all datasets from the dissimilar experimental artifacts; (*ii*) one of multiple types of genomic information, such as DNA copy number, DNA methylation and mRNA expression, collected from the same set of samples, e.g., tumor samples, to elucidate the molecular composition of the overall biological signal in these samples; (*iii*) one of multiple chromosomes of the same organism, to illustrate the relation, if any, between these chromosomes in terms of their, e.g., mRNA expression in a given set of samples; and (*iv*) one of multiple interacting organisms, e.g., in an ecosystem, to illuminate the exchange of biological information in these interactions.

Supporting Information

Appendix S1 A PDF format file, readable by Adobe Acrobat Reader.
(PDF)

Mathematica Notebook S1 Higher-order generalized singular value decomposition (HO GSVD) of global mRNA expression datasets from three different organisms. A Mathematica 5.2 code file, executable by Mathematica 5.2 and readable by Mathematica Player, freely available at http://www.wolfram.com/products/player/.
(NB)

Mathematica Notebook S2 HO GSVD of global mRNA expression datasets from three different organisms. A PDF format file, readable by Adobe Acrobat Reader.
(PDF)

Dataset S1 *S. pombe* global mRNA expression. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, reproducing the relative mRNA expression levels of $m_1 = 3167$ *S. pombe* gene clones at $n = 17$ time points during about two cell-cycle periods from Rustici *et al.* [15] with the cell-cycle classifications of Rustici *et al.* or Oliva *et al.* [16].
(TXT)

Dataset S2 *S. cerevisiae* global mRNA expression. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, reproducing the relative mRNA expression levels of $m_2 = 4772$ *S. cerevisiae* open reading frames (ORFs), or genes, at $n = 17$ time points during about two cell-cycle periods, including cell-cycle classifications, from Spellman *et al.* [17].
(TXT)

Dataset S3 Human global mRNA expression. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, reproducing the relative mRNA expression levels of $m_3 = 13{,}068$ human genes at $n = 17$ time points during about two cell-cycle periods, including cell-cycle classifications, from Whitfield *et al.* [18].
(TXT)

Acknowledgments

We thank G. H. Golub for introducing us to matrix and tensor computations, and the American Institute of Mathematics in Palo Alto and Stanford University for hosting the 2004 Workshop on Tensor Decompositions and the 2006 Workshop on Algorithms for Modern Massive Data Sets, respectively, where some of this work was done. We also thank C. H. Lee for technical assistance, R. A. Horn for helpful discussions of matrix analysis and careful reading of the manuscript, and L. De Lathauwer and A. Goffeau for helpful comments.

Author Contributions

Conceived and designed the experiments: OA. Performed the experiments: SPP OA. Analyzed the data: SPP OA. Contributed reagents/materials/analysis tools: SPP OA. Wrote the paper: SPP MAS CFVL OA. Proved mathematical theorems: SPP MAS CFVL OA.

References

1. Jensen LJ, Jensen TS, de Lichtenberg U, Brunak S, Bork P (2006) Co-evolution of transcriptional and post-translational cell-cycle regulation. Nature 443: 594–597.
2. Lu Y, Huggins P, Bar-Joseph Z (2009) Cross species analysis of microarray expression data. Bioinformatics 25: 1476–1483.
3. Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci USA 93: 10268–10273.
4. Golub GH, Van Loan CF (1996) Matrix Computations. Baltimore: Johns Hopkins University Press, third edition. 694 p.
5. Van Loan CF (1976) Generalizing the singular value decomposition. SIAM J Numer Anal 13: 76–83.
6. Paige CC, Saunders MA (1981) Towards a generalized singular value decomposition. SIAM J Numer Anal 18: 398–405.
7. Alter O, Brown PO, Botstein D (2003) Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. Proc Natl Acad Sci USA 100: 3351–3356.
8. Alter O (2006) Discovery of principles of nature from mathematical modeling of DNA microarray data. Proc Natl Acad Sci USA 103: 16063–16064.
9. Omberg L, Meyerson JR, Kobayashi K, Drury LS, Diffley JF, et al. (2009) Global effects of DNA replication and DNA replication origin activity on eukaryotic gene expression. Mol Syst Biol 5: 312.
10. Alter O, Golub GH (2004) Integrative analysis of genome-scale data by using pseudoinverse projection predicts novel correlation between DNA replication and RNA transcription. Proc Natl Acad Sci USA 101: 16577–16582.
11. Omberg L, Golub GH, Alter O (2007) A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. Proc Natl Acad Sci USA 104: 18371–18376.
12. De Lathauwer L, De Moor B, Vandewalle J (2000) A multilinear singular value decomposition. SIAM J Matrix Anal Appl 21: 1253–1278.
13. Vandewalle J, De Lathauwer L, Comon P (2003) The generalized higher order singular value decomposition and the oriented signal-to-signal ratios of pairs of signal tensors and their use in signal processing. In: Proc ECCTD'03 - European Conf on Circuit Theory and Design. pp I-389–I-392.
14. Alter O, Golub GH (2005) Reconstructing the pathways of a cellular system from genome-scale signals using matrix and tensor computations. Proc Natl Acad Sci USA 102: 17559–17564.
15. Rustici G, Mata J, Kivinen K, Lió P, Penkett CJ, et al. (2004) Periodic gene expression program of the fission yeast cell cycle. Nat Genet 36: 809–817.
16. Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, et al. (2005) The cell cycle-regulated genes of Schizosaccharomyces pombe. PLoS Biol 3: e225.
17. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 9: 3273–3297.
18. Whitfield ML, Sherlock G, Saldanha A, Murray JI, Ball CA, et al. (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13: 1977–2000.
19. Gauthier NP, Larsen ME, Wernersson R, Brunak S, Jensen TS (2010) Cyclebase.org: version 2.0, an updated comprehensive, multi-species repository of cell cycle experiments and derived analysis results. Nucleic Acids Res 38: D699–D702.
20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
21. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated nonredundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35: D61–D65.
22. Decottignies A, Goffeau A (1997) Complete inventory of the yeast ABC proteins. Nat Genet 15: 137–145.
23. Mamnun YM, Schüller C, Kuchler K (2004) Expression regulation of the yeast PDR5 ATP-binding cassette (ABC) transporter suggests a role in cellular detoxification during the exponential growth phase. FEBS Lett 559: 111–117.
24. Werner-Washburne M, Braun E, Johnston GC, Singer RA (1993) Stationary phase in the yeast Saccharomyces cerevisiae. Microbiol Rev 57: 383–401.
25. Meyers S, Schauer W, Balzi E, Wagner M, Goffeau A, et al. (1992) Interaction of the yeast pleiotropic drug resistance genes PDR1 and PDR5. Curr Genet 21: 431–436.
26. Mahé Y, Parle-McDermott A, Nourani A, Delahodde A, Lamprecht A, et al. (1996) The ATP-binding cassette multidrug transporter Snq2 of Saccharomyces cerevisiae: a novel target for the transcription factors Pdr1 and Pdr3. Mol Microbiol 20: 109–117.
27. Wolfger H, Mahé Y, Parle-McDermott A, Delahodde A, Kuchler K (1997) The yeast ATP binding cassette (ABC) protein genes PDR10 and PDR15 are novel targets for the Pdr1 and Pdr3 transcriptional regulators. FEBS Lett 418: 269–274.
28. Hlaváček O, Kučerová H, Harant K, Palková Z, Váchová L (2009) Putative role for ABC multidrug exporters in yeast quorum sensing. FEBS Lett 583: 1107–1113.
29. Lee KS, Patton JL, Fido M, Hines LK, Kohlwein SD, et al. (1994) The Saccharomyces cerevisiae PLB1 gene encodes a protein required for lysophospholipase and phospholipase B activity. J Biol Chem 269: 19725–19730.
30. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, et al. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2: 65–73.
31. Martin-Castellanos C, Labib K, Moreno S (1996) B-type cyclins regulate G1 progression in fission yeast in opposition to the $p25^{rum1}$ cdk inhibitor. EMBO J 15: 839–849.

32. Fisher DL, Nurse P (1996) A single fission yeast mitotic cyclin B $p34^{cdc2}$ kinase promotes both S-phase and mitosis in the absence of G1 cyclins. EMBO J 15: 850–860.
33. Chu MT, Funderlic RE, Golub GH (1997) On a variational formulation of the generalized singular value decomposition. SIAM J Matrix Anal Appl 18: 1082–1092.
34. Rao CR (1973) Linear Statistical Inference and Its Applications. New York, NY: John Wiley & Sons, second edition. 656 p.
35. Rao CR (1984) Optimization of functions of matrices with applications to statistical problems. In: Rao PSRS, Sedransk J, eds. W.G. Cochran's Impact on Statistics. New York, NY: John Wiley & Sons. pp 191–202.
36. Paige CC (1985) The general linear model and the generalized singular value decomposition. Linear Algebra Appl 70: 269–284.
37. Marshall AW, Olkin L (1990) Matrix versions of the Cauchy and Kantorovich inequalities. Aequationes Mathematicae 40: 89–93.
38. Horn RA, Johnson CR (1985) Matrix Analysis. Cambridge, UK: Cambridge University Press. 575 p.
39. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. Nat Genet 22: 281–285.